Teoria dos *rough sets* com mapas auto-organizáveis de *kohonen* na descoberta de conhecimento em bases de dados

Renato José Sassi
rjsassi@gmail.com

RESUMO
As bases de dados do mundo real contêm grande volume de dados, e entre eles estão escondidas relações interessantes que são realmente muito difíceis de descobrir. Assim, os Sistemas de Descoberta de Conhecimento em Bases de Dados (KDD – Knowledge Discovery in Databases) surgem como uma possível solução para descobrir essas relações com o objetivo de converter informação em conhecimento. No entanto, nem todos esses dados são úteis para o KDD. Em geral, são pré-processados antes de ser apresentados ao KDD, o que visa diminuir a quantidade e também selecionar os dados mais relevantes a serem utilizados pelo KDD. Este trabalho consiste na utilização da Teoria dos Rough Sets, a fim de pré-processar dados que foram apresentados a uma rede neural artificial do tipo Mapas Auto-Organizáveis de Kohonen ou Self-Organizing Maps (rede SOM), formando uma arquitetura híbrida. Os resultados experimentais indicaram um melhor desempenho da arquitetura híbrida em relação a uma rede neural artificial Mapas Auto-Organizáveis de Kohonem. O trabalho apresenta também todas as fases do processo de KDD.

Palavras-chave: *Arquitetura híbrida, Redes neurais, Descoberta de conhecimento, Teoria dos rough sets, Mapas auto-organizáveis.*

Rough sets theory with self-organizing maps of *kohonen* in knowledge discovery in databases

ABSTRACT
The databases of real world contains a huge volume of data and among them there are hidden piles of interesting relations that are actually very hard to find out. The knowledge discovery databases (KDD) appear as a possible solution to find out such relations aiming at converting information into knowledge. However, not all data presented in the bases are useful to a KDD. Usually, data are processed before being presented to a KDD aiming at reducing the amount of data and also at selecting more relevant data to be used by the system. This work consists in the use of Rough Sets Theory, in order to pre-processing data to be presented to Self-Organizing Maps neural network (hybrid architecture). Experiments' results evidence the better performance using the hybrid architecture than Self-Organizing Maps. The paper also presents all phases of the KDD process.

Keywords: *Hybrid architecture, Neural network, Knowledge discovery systems, Rough sets, Self-organizing maps.*

1. Introdução

Os avanços na área da Tecnologia da Informação têm possibilitado o armazenamento de grandes e múltiplas bases de dados. Esses dados produzidos e armazenados em larga escala são difíceis de serem analisados, interpretados e relacionados pelos métodos tradicionais, como planilhas de cálculo e relatórios informativos operacionais. Por essa razão, faz-se necessário o uso de sistemas que possam extrair o conhecimento dessas bases, viabilizando a análise dos dados, denominados KDD (*Knowledge Discovery in Databases*) ou Descoberta de Conhecimento em Bases de Dados.

O KDD pode ser definido como um processo de extração de conhecimentos válidos, novos, potencialmente úteis e compreensíveis para apoiar a tomada de decisão (FAYYAD et al., 1996). Para tanto, o KDD utiliza as seguintes áreas para realizar os seus processos: métodos estatísticos, reconhecimento de padrões, visualização, banco de dados, aprendizado de máquina, Inteligência Artificial e *Data Warehouse*, entre outras.

O KDD é um processo constituído de fases que possuem inúmeros passos, os quais envolvem número elevado de decisões a serem tomadas pelo usuário, ou seja, é um processo interativo. É também processo iterativo, pois, ao longo do processo de KDD, um passo será repetido tantas vezes quantas se fizerem necessárias para que se chegue a um resultado satisfatório. Entretanto, nem todos os dados que compõem as bases servem para um sistema descobrir conhecimento. Assim, é necessário pré-processar os dados antes de seguir com o processo de descoberta de conhecimento.

A diminuição de dados é uma forma de pré-processamento que visa obter uma representação reduzida dos dados, mas que produz os mesmos (ou quase os mesmos) resultados analíticos.

A Teoria dos *Rough Sets* (RS), ou Teoria dos Conjuntos Aproximados (TCA), foi proposta por Pawlak (1982) como um novo modelo matemático para representação do conhecimento e tratamento de incerteza. Devido a suas características, a Teoria dos *Rough Sets* é muito empregada na redução de dados.

As redes neurais artificiais são redes inspiradas na estrutura do cérebro, com o objetivo de apresentar características similares ao do comportamento humano, como: aprendizado, associação, generalização e abstração.

Uma rede neural artificial é um processador maciçamente paralelo, distribuído, constituído de unidades de processamento simples que tem capacidade para armazenar conhecimento experimental e torná-lo disponível para uso (HAYKIN, 1999).

As redes neurais artificiais têm sido amplamente utilizadas nas mais variadas aplicações, incluindo mineração de dados (SASSI et al., 2008). A rede SOM é uma das diversas arquiteturas de redes neurais artificiais e foi escolhida neste trabalho porque, no caso de KDD (VESANTO; ALHONIEMI, 2000), possibilita em um mapa bidimensional a formação e visualização simples dos *clusters* (grupos) e da correlação dos dados, preservando a posição relativa desses *clusters* no hiperespaço original, ou seja, é utilizada para a tarefa de clusterização. Entretanto, uma das desvantagens da rede SOM é a imprecisão na definição de fronteira entre os *clusters* (LABIOD et al., 2010).

Existem outras técnicas aplicadas na tarefa de clusterização, Kumar e Dhamija (2010) fazem uma análise comparativa entre a rede SOM e o algoritmo de clusterização K-*means*.

Segundo Goldschmidt e Passos (2005), para suprir desvantagens como as citadas anteriormente, técnicas podem ser combinadas para gerar as chamadas arquiteturas híbridas. A grande vantagem desse tipo de sistema deve-se ao sinergismo obtido pela combinação de duas ou mais técnicas. Esse sinergismo reflete na obtenção de um sistema mais poderoso e com menos deficiências. Assim, a motivação deste trabalho reside em desenvolver uma arquitetura híbrida que combina a Teoria dos *Rough Sets* com uma rede neural artificial tipo Mapas Auto-Organizáveis de Kohonen ou rede SOM (*Self-Organizing Maps*) (KOHONEN, 2001).

A arquitetura híbrida funciona da seguinte forma: a base de dados é apresentada ao *Rough Sets* realizando a redução dos atributos que provocam incerteza, e, posteriormente, a base reduzida é apresentada à rede SOM para geração de *clusters* (agrupamentos). O objetivo com a arquitetura híbrida é eliminar informação desnecessária que possa ser apresentada à rede, aumentando a sua capacidade de gerar *clusters* mais coesos.

A fim de verificar o sinergismo, comparou-se a arquitetura híbrida (RS + rede SOM) com uma rede SOM sem a presença de *Rough Sets*. Os resultados foram avaliados com base na resolução do mapa e em relação à preservação da topologia dos dados de entrada (KIVILUOTO, 1996). O melhor resultado deve ser aquele que "melhor representa os dados de entrada". Esse critério normalmente é traduzido por duas medidas: o erro de quantização (EQ) e o erro topográfico (ET). Além das medidas de qualidade, foram adotados também como critério de avaliação: o número de *clusters* gerados e a visualização através do Mapa por Similaridade de Cor.

Outros trabalhos associaram RS à rede SOM em uma arquitetura híbrida (PAL et al., 2004; MOHEBI; SAP, 2010), utilizando o conceito de espaços aproximados dos RS para melhorar a formação dos *clusters* pela rede SOM.

O processo de KDD apresenta melhor resultado quando submetido à análise de grandes bases de dados (CASTANHEIRA, 2009). No caso do trabalho proposto, a base de dados escolhida não é extensa em número de registros, mas possui bom número de atributos (48), o que interessa ao RS, pois a técnica reduz atributos e não registros.

Além da introdução, a sequência do trabalho é organizada da seguinte forma: na seção 2, os principais conceitos do KDD são discutidos; na seção 3, conceitos principais da Teoria dos *Rough Sets* são apresentados - Uma explanação resumida da rede SOM e das principais medidas de desempenho é discutida na seção 4. A metodologia e os experimentos realizados são apresentados na seção 5 e na seção 6, os resultados. Na última seção, conclui-se o trabalho.

2. Descoberta de conhecimento em bases de dados (KDD)

KDD é o termo criado pelo Gartner *Group* na década de 1980 para descrever todo o processo de extração de conhecimento dos dados devido ao crescimento vertiginoso das bases de dados das organizações.

O KDD é um processo composto por fases que devem ser desenvolvidas para atingir o objetivo final, que é a extração de conhecimento. As fases do KDD possuem numerosos passos, que envolvem número elevado de decisões a serem tomadas, ou seja, é um processo interativo. É, também, um processo iterativo, pois, ao longo do processo KDD, um passo será repetido tantas vezes quantas se fizerem necessárias para que se chegue a um resultado satisfatório (FAYYAD *et al*., 1996).

No processo de KDD, cada fase possui intersecção com as demais. Desse modo, os resultados produzidos numa fase são utilizados para melhorar os resultados das próximas. O KDD é composto das seguintes fases: seleção dos dados, pré-processamento dos dados, transformação dos dados, mineração de dados (*Data Mining*) e interpretação/avaliação do conhecimento. A Figura 1 ilustra as fases do processo do KDD.

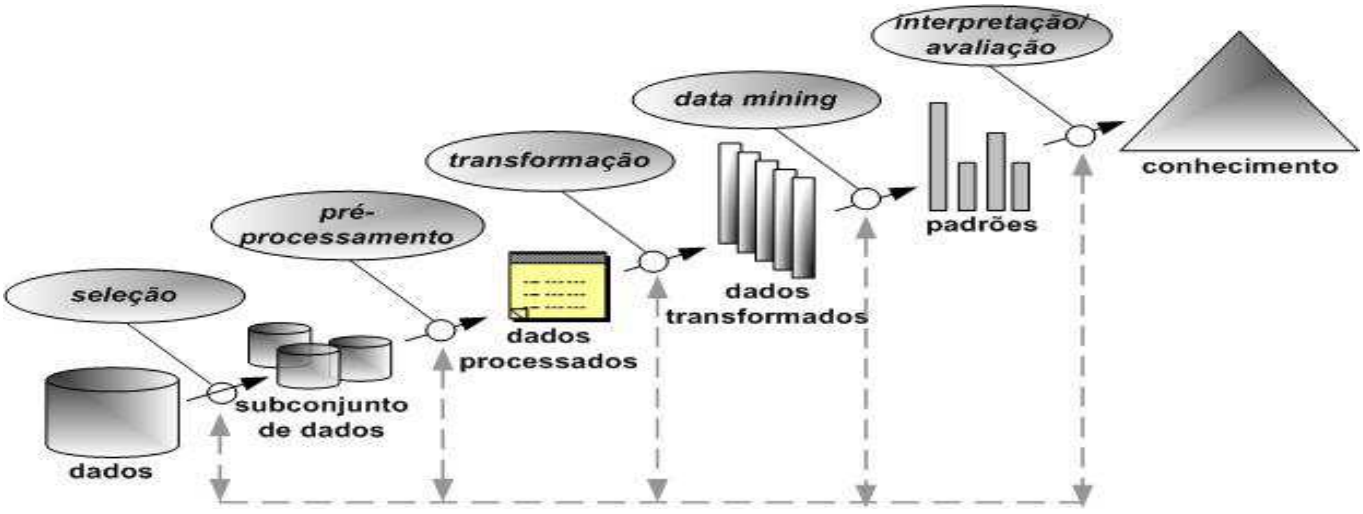

Figura 1 - Fases do processo de KDD. A iteração entre as fases pode ser observada pelas setas tracejadas
Fonte: Adaptada de FAYYAD *et al*., 1996.

As três fases iniciais do KDD (Figura 1), que envolvem a seleção, o pré-processamento e a transformação, também chamadas de preparação dos dados, exigem bastante tempo, aproximadamente entre 60 e 80% do tempo utilizado em todo o processo, sendo a maior parte desse tempo consumida com a limpeza dos dados. O foco será dado no pré-processamento, no *Data Mining* e na Interpretação/Avaliação do Conhecimento.

O pré-processamento dos dados tem por objetivo assegurar a qualidade dos dados selecionados.

A limpeza dos dados envolve a verificação da consistência das informações, a correção de possíveis erros e o preenchimento ou a eliminação de valores nulos e redundantes. Nessa fase são identificados e removidos os dados, duplicados e, ou, corrompidos. A limpeza dos dados consiste em resolver problemas com: dados com erros (valores discrepantes ou *outliers*), registros repetidos e valores faltantes.

A redução de dados é considerada técnica de pré-processamento dos dados, e seus estudos têm importância neste trabalho, porque a Arquitetura Híbrida proposta combina o *Rough Sets* como técnica de redução de atributos para a rede SOM.

O termo *Data Mining* surgiu devido às semelhanças entre a procura de informação importante para o mundo dos negócios (numa base de dados) e o ato de minerar a montanha para encontrar um veio de ouro.

A mineração de dados é considerada a etapa mais importante do processo de KDD. Caracteriza-se pela existência do algoritmo minerador (*Data Mining*), que diante da tarefa especificada será capaz de extrair, de modo eficiente, conhecimento implícito e útil de um banco de dados.

Segundo Berry e Linoff (1997), *Data Mining* é a exploração e análise, por meios automáticos ou semiautomáticos, de grandes quantidades de dados para descobrir modelos e regras significativas, permitindo a uma empresa aumentar, por exemplo, suas operações de marketing, vendas e apoio aos clientes pela melhor compreensão da clientela. De forma mais simples, *Data Mining* é produzir conhecimento novo escondido em grandes bases de dados.

O *Data Mining* usa técnicas baseadas em descobertas por meio de procura de padrões dos dados, o que é feito com o emprego de algoritmos inteligentes para encontrar relações fundamentais entre os dados.

As técnicas de *Data Mining* permitem avaliar como as perguntas se relacionam com as respostas (padrões e relações) encontradas. Achados essas relações e padrões, é fornecida uma base de regras que servem de apoio aos processos de tomada de decisão. Para tal, utilizam-se técnicas baseadas em Inteligência Artificial, como as Redes Neurais Artificiais, as Árvores de Decisões, a Teoria dos Conjuntos *Fuzzy*, os Algoritmos Genéticos ou, ainda, combinações entre essas técnicas, gerando as chamadas Arquiteturas Híbridas (HAN; KAMBER, 2001).

O resultado obtido pela aplicação do *Data Mining* deve ser compacto, legível (apresentado de alguma forma simbólica) e interpretável e representar fielmente os dados que lhe deram origem.

Uma questão fundamental do *Data Mining* é quanto à qualidade do conhecimento extraído, levando em consideração a precisão, a compreensibilidade, a surpresa e, ou, o potencial interesse do conhecimento obtido. É necessário verificar o que foi aprendido, o que há de novo, eliminando o conhecimento "inútil" e muito óbvio. Mongiovi (1995) exemplificou: se uma cadeia de lojas resolve buscar associações entre os itens que ela vende, certamente haverá associações entre itens que farão sentido, e outras que não farão. Faz-se necessário, então, que um especialista dos negócios da empresa avalie quais são as associações realmente relevantes para a empresa. Por último, as associações descobertas precisam ser acionáveis, ou seja, é necessário que possam ser realizadas ações simples para que o conhecimento gerado seja traduzido em vantagem aos negócios da empresa.

O *Data Mining* pode responder a questões de negócio que tradicionalmente demandariam muito tempo para resolver. Ele explora as bases de dados à procura de padrões escondidos, encontrando dados que permitem prever tendências e comportamentos futuros, que os especialistas podem não descobrir devido ao fato de essa informação sair do limite de suas expectativas, possibilitando a tomada de decisão.

Após a fase de mineração de dados é necessária a interpretação do conhecimento descoberto, ou algum processamento desse

conhecimento. Em geral, a principal meta dessa fase é melhorar a compreensão do conhecimento descoberto pelo algoritmo minerador, validando-o através de medidas da qualidade da solução e da percepção de um analista de dados. Esses conhecimentos serão consolidados em forma de relatórios demonstrativos, com a documentação e explicação das informações relevantes ocorridas em cada etapa do processo de KDD. Uma das maneiras genéricas de obter a compreensão e interpretação dos resultados é utilizar técnicas de visualização.

2.1. Tarefas do KDD

Existem várias formas de interpretação dos dados pelo KDD denominadas tarefas. As tarefas mais comuns são (FAYYAD *et al.*, 1996): associação, classificação, clusterização (agrupamento) e visualização. O foco será dado nas tarefas de clusterização e visualização que são aplicadas neste trabalho.

A clusterização transforma registros com grande número de atributos em conjuntos relativamente menores (segmentos). Essa transformação é realizada, automaticamente, por meio de identificação das características que distinguem o conjunto de dados e pelo seu posterior particionamento (HAN; KAMBER, 2001). Não é necessário identificar os agrupamentos desejados nem os atributos que devem ser utilizados para a produção dos segmentos. O objetivo nessa tarefa é maximizar similaridade intra-*cluster* e minimizar similaridade extra-*cluster* (GOLDSCHMIDT; PASSOS, 2005).

Os resultados de uma operação de clusterização podem ser usados de duas diferentes maneiras: ora para produzir um sumário da base de dados, por meio das características de cada *cluster*, ora como dados de entrada para outras técnicas, por exemplo, a classificação.

A clusterização pode ser usada em casos que façam uso de modelos de segmentação de população, como segmentação demográfica de mercados de consumidores (identificar grupos homogêneos de elementos, identificar elementos dentro do mesmo grupo maximamente semelhantes), implicando possível comparação dos hábitos de consumo de múltiplos segmentos de população, visando determinar campanhas de vendas.

As ferramentas de visualização não são propriamente tarefas de *Data Mining*, mas sim meios de analisar e observar os dados de determinada base de dados (BERRY; LINOFF, 1997).

A visualização fornece meios de obter sumários visuais dos dados de uma base de dados. No caso de técnicas de clusterização, podem ser usadas ferramentas de visualização para determinar quais *clusters* são úteis ou interessantes para as técnicas de *Data Mining*. No caso específico da rede SOM, as formas de visualização mais utilizadas são a Matriz-U (ULTSCH, 1995) e o Mapa por Similaridade de Cor.

As ferramentas de visualização podem, ainda, ser usadas como mecanismo de compreensão da informação extraída por meio das técnicas de *Data Mining*. Características difíceis de detectar pela simples observação de linhas e colunas com valores numéricos podem se tornar óbvias se forem observadas graficamente. Por meio dessas ferramentas podem ser encontrados características ou fenômenos pouco comuns ou interessantes sem que se esteja diretamente procurando por eles.

3. Teoria dos *Rough Sets*

A Teoria dos *Rough Sets* (RS) foi proposta por Pawlak (1982) como um novo modelo matemático para representação do conhecimento e tratamento de incerteza, tendo sido usada, posteriormente, para o desenvolvimento de técnicas para classificação aproximada em aprendizado de máquina. Devido a essas características, tem-se utilizado essa teoria em Inteligência Artificial, especialmente nas áreas de aquisição de conhecimento, raciocínio indutivo e descoberta de conhecimento em base de dados.

Conjuntos aproximados podem ser considerados conjuntos com fronteiras nebulosas, ou seja, conjuntos que não podem ser caracterizados precisamente, utilizando-se dos atributos disponíveis (PAWLAK, 1991).

A incerteza pode-se manifestar de diversas formas, como: imprecisão, incompletude, inconsistência etc. RS trata de um tipo fundamental de incerteza, a indiscernibilidade. A indiscernibilidade surge quando não é possível distinguir elementos de um mesmo conjunto, e representa a situação em que esses elementos parecem todos ser um único elemento (UCHOA, 1998).

Os conceitos de RS têm-se mostrado muito úteis quando aplicados a problemas do tipo: redução de atributos, descoberta de dependência entre atributos e descoberta de padrões entre os dados (PAWLAK, 1991).

A diminuição de atributos realizada pelos RS é feita através dos chamados de redutos, que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais. Tal procedimento de eliminação de atributos irrelevantes é uma das características da Teoria. Uma interessante aplicação dos RS em atributos pode ser verificada em Sant'anna (2008).

3.1. Espaços aproximados

Um espaço aproximado é um par ordenado *A* = (U,R), em que U é um conjunto não vazio, denominado conjunto universo; e R é uma relação de equivalência sobre U, denominada Relação de Indiscernibilidade. Uma relação binária $R \subseteq X \times X$, a qual é reflexiva (um elemento está relacionado com ele próprio xRx), simétrica (se xRy então yRx) e transitiva (se xRy e yRz então xRz), é chamada de relação de equivalência. Dados os elementos x, y ∈ U, se xRy então x e y são indiscerníveis em *A*,ou seja, a classe de equivalência definida por x é a mesma que a definida por y, i.e., [x]R = [y]R.

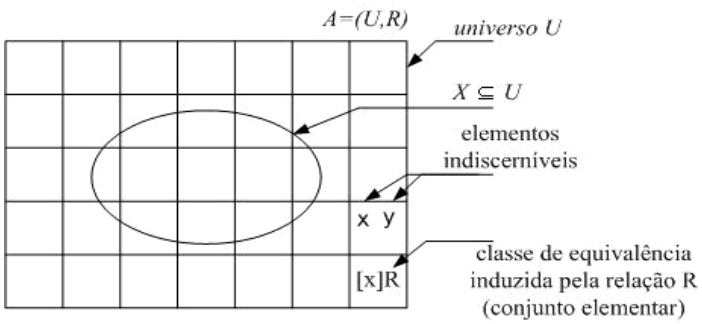

Figura 2 - Espaço aproximado *A* = (U,R) e X ⊆ U
Fonte: Adaptado de PAWLAK, 1982; 1991.

A classe de equivalência de um elemento x ∈ X consiste de todos os elementos y ∈ X para os quais xRy. Os elementos que são indiscerníveis formam conjuntos chamados de conjuntos elementares. Dessa forma, pode-se dizer que as classes de equivalência de R são os conjuntos elementares de *A.* Na Figura 2, pode-se visualizar o espaço aproximado *A* = (U,R).

Os principais conceitos dos RS são: Espaços Aproximados, Aproximação Inferior (AI), Aproximação Superior (AS), Sistema de Informação (S), Sistema de Decisão (SD) e Indiscernibilidade (IND).

Este trabalho não tem como finalidade o aprofundamento no formalismo matemático dos RS, que é grande. Para isso, recomenda-se o trabalho de Uchôa (1998).

A forma mais comum para representação dos dados em RS é através de um Sistema de Informação (S) (Tabela 1) que contém um conjunto de elementos, e cada elemento tem quantidade de atributos condicionais. Esses atributos são os mesmos para cada um dos elementos, mas os seus valores nominais podem diferir (Tabela 1).

Dessa forma, um Sistema de Informação é um par ordenado S = (U, C), em que U é um conjunto finito e não vazio de elementos chamado de universo (Figura 2), e C é um conjunto finito e não vazio formado pelos atributos. Cada atributo **a**∈ C é uma função **a**: U → V**a**, em que V**a** é o conjunto dos valores permitidos para o atributo **a** (sua faixa de valores). Na Tabela 1, em que é apresentado o Sistema de Informação S, podem-se observar os principais conceitos de RS, o espaço aproximado *A* = (U,R), o universo U formado pelos elementos e1, e2, e3, e4, e5 e e6 e os atributos (C) Experiência do Vendedor, Qualidade do Produto e Boa Localização e R a relação de equivalência sobre U.

3.2. Indiscernibilidade

O principal conceito envolvido em RS é a Relação de Indiscernibilidade (PAWLAK, 1982), a qual normalmente está associada a um conjunto de atributos. Se tal relação existe entre dois elementos, isso significa que todos os valores nominais dos seus atributos são idênticos com respeito aos atributos considerados, portanto não podem ser discernidos (distinguidos) entre si.

Para cada subconjunto de atributos B ⊆ C no Sistema de Informação S = (U, C) é associada uma relação de equivalência INDs(B), chamada de Relação de Indiscernibilidade, definida como: $IND_S(B) = \{(x,y) \in U^2 / \forall \mathbf{a} \in B, \mathbf{a}(x) = \mathbf{a}(y)\}$. O conjunto de todas as classes de equivalência na relação INDs(B) é representado por U/INDs(B), denominado quociente de U pela relação INDs(B).

Tabela 1 - Exemplo de um Sistema de Informação (S)

LOJA	Experiência do Vendedor	Qualidade do Produto	Boa Localização
e1	Alta	Boa	Não
e2	Média	Boa	Não
e3	Média	Boa	Não
e4	Baixa	Média	Não
e5	Média	Média	Sim
e6	Alta	Média	Sim

Fonte: Adaptado de PAWLAK, 1991.

Em muitos casos, é importante a classificação dos elementos considerando um atributo de decisão que informa a decisão a ser tomada. Assim, um S que apresenta atributo de decisão é denominado Sistema de Decisão (SD). Um SD pode ser representado por SD = (U, C ∪ {d}), em que d ∉ C é o atributo de decisão. A Tabela 2 mostra um SD obtido a partir do Sistema de Informação S da Tabela 1, destacando-se os atributos condicionais (Experiência do Vendedor, Qualidade do Produto e Boa Localização) e o atributo de decisão (Retorno).

Tabela 2 - Sistema de decisão (Sistema de Informação com o atributo de decisão Retorno)

	Atributos Condicionais			Atributo de Decisão
Loja	**Experiência do Vendedor**	**Qualidade do Produto**	**Boa Locali-zação**	**Retorno**
e1	Alta	Boa	Não	Lucro
e2	Média	Boa	Não	Prejuízo
e3	Média	Boa	Não	Lucro
e4	Baixa	Média	Não	Prejuízo
e5	Média	Média	Sim	Prejuízo
e6	Alta	Média	Sim	Lucro

Fonte: Adaptado de PAWLAK, 1991.

Os valores dos atributos são chamados de valores nominais e estão expressos como: Experiência do Vendedor {Alta, Média, Baixa}; Qualidade do Produto {Boa, Média}; Boa Localização {Não, Sim}; e Retorno {Lucro, Prejuízo}. Considerando cada atributo condicional de forma independente, a relação de equivalência do sistema de informação S (Tabela 2) forma os seguintes conjuntos elementares: experiência do vendedor Alta {e1, e6}; Média {e2, e3, e5}; Baixa {e4}; Qualidade do Produto: Boa {e1, e2, e3}; Média {e4, e5, e6} e Boa Localização: Não {e1, e2, e3, e4}; e Sim {e5, e6}.

Ao utilizar todos os atributos condicionais do Sistema de Informação S da Tabela 1, obtêm-se os seguintes conjuntos elementares: {e1}, {e2, e3}, {e4}, {e5} e {e6}. Observando a Tabela 3, pode-se perceber que há dois elementos (casos) {e2} e {e3} iguais (destacados em negrito), no que se refere a valores de atributos condicionais.

Existindo a Relação de Indiscernibilidade entre os elementos {e2} e {e3} como mostrado na Tabela 3, isso significa que todos os valores nominais de seus atributos são idênticos com relação ao subconjunto de atributos B (B ⊆ S) considerado, ou seja, não podem ser diferenciados entre si.

Existindo a Relação de Indiscernibilidade entre os elementos {e2} e {e3} como mostrado na Tabela 3, isso significa que todos os valores nominais de seus atributos são idênticos com relação ao subconjunto de atributos B (B ⊆ S) considerado, ou seja, não podem ser diferenciados entre si.

Parte interessante da Teoria é a Aproximação de Conjuntos, que utiliza os conceitos de Aproximação Inferior, Aproximação Superior e Fronteira, os quais não foram abordados, pois, neste trabalho, a aplicação dos RS sets se resume à redução de atributos.

Tabela 3 - Sistema de Decisão com os elementos e2 e e3 indiscerníveis, com relação aos atributos condicionais

Loja	Experiência do Vendedor	Qualidade do Produto	Boa Localização	Retorno
e1	Alta	Boa	Não	Lucro
e2	**Média**	**Boa**	**Não**	**Prejuízo**
e3	**Média**	**Boa**	**Não**	**Lucro**
e4	Baixa	Média	Não	Prejuízo
e5	Média	Média	Sim	Prejuízo
e6	Alta	Média	Sim	Lucro

Fonte: Adaptado de PAWLAK, 1991.

3.3 Redução de Atributos

A redução de atributos em RS é feita através dos chamados Redutos (RED), que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais (PAWLAK, 1982).

Um Reduto de B sobre um sistema de informação S é um conjunto de atributos B' $\subseteq$ B tal que todos os atributos **a**$\in$ (B – B') são dispensáveis. Com isso, U/INDs(B') = U/INDs(B). O termo RED(B) é utilizado para denotar a família de redutos de B.

O cálculo de reduções para gerar os redutos é um problema *n-p* completo, e seu processamento em grandes bases de dados exige grande esforço computacional.

Essa redução é feita pela função de discernibilidade, a partir da Matriz de Discernibilidade. Considerando o conjunto de atributos B = {Experiência do Vendedor, Qualidade do Produto e Boa Localização} no Sistema de Informação S, o conjunto de todas as classes de equivalência determinadas por B sobre S é dado por U/INDs(B) = {{e1}; {e2, e3}; {e4}; {e5}; {e6}}, que estão representadas na Tabela 3.

A Matriz de Discernibilidade do Sistema de Informação S, denotada por MD(B), é uma matriz simétrica n x n com: mD(i, j) = {**a**$\in$ B $|$**a**(Ei) $\neq$**a**(Ej)} para i,j = 1, 2, ...,n.. sendo 1$\leq$i, j$\leq$n e n=$|$ U / INDs(B). Logo, os elementos da Matriz de Discernibilidade mD(i, j) é o conjunto de atributos condicionais de B que diferenciam os elementos das classes com relação aos seus valores nominais. Considerando Experiência do Vendedor (EV), Qualidade do Produto (QP) e Boa Localização (BL), com a finalidade de construir a Matriz de Discernibilidade MD(B), tem-se na Tabela 4 a sua representação.

Tabela 4 - Matriz de Discernibilidade

	e1	e2	e3	e4	e5	e6
e1	∅					
e2	EV	∅				
e3	EV	∅	∅			
e4	EV, QP	EV, QP	EV, QP	∅		
e5	EV, QP, BL	QP, BL	QP, BL	EV, BL	∅	
e6	QP, BL	EV, QP, BL	EV, QP, BL	EV, BL	EV	∅

A função de discernibilidade Fs(B) é uma função booleana com *m* variáveis, que determina o conjunto mínimo de atributos necessários para diferenciar qualquer classe de equivalência das demais, definida como:

$$F_S(a_1^*, a_2^*, \ldots, a_m^*) = \wedge\left\{\vee m_D^*(i,j) \mid i, j = 1,2,\ldots,n, \quad m_D(i,j) \neq 0\right\}$$

Sendo: $m_D^*(i,j) = \left\{\mathbf{a}^* \mid \mathbf{a} \in m_D(i,j)\right\}$

Utilizando o método de simplificação de expressões booleanas na função Fs(B), obtém-se o conjunto de todos os implicantes primos dessa função, o qual determina os redutos de S.

A simplificação é um processo de manipulação algébrica das funções lógicas com a finalidade de diminuir o número de variáveis e de operações necessárias para a sua realização (PATRICIO et al., 2005).

A função de discernibilidade Fs(B) é obtida da seguinte forma: para os atributos contidos dentro de cada célula da Matriz de Discernibilidade MD(B) (Tabela 4), aplica-se o operador "soma", "*or*" ou "$\vee$" e, entre as células dessa matriz, utiliza-se o operador "produto", "*and*" ou "$\wedge$", resultando em uma expressão booleana de "Produto da Soma". A Fs(B) da Tabela 4 é representada por:

Fs(B) = (EV) $\wedge$ (EV) $\wedge$ (EV $\vee$ QP) $\wedge$ (EV $\vee$ QP) $\wedge$ (EV $\vee$ QP) $\wedge$ (EV $\vee$ QP $\vee$ BL) $\wedge$ (QP $\vee$ BL) $\wedge$ (QP $\vee$ BL) $\wedge$ (EV $\vee$ BL) $\wedge$ (QP $\vee$ BL) $\wedge$ (EV $\vee$ QP $\vee$ BL) $\wedge$ (EV $\vee$ QP $\vee$ BL) $\wedge$ (EV $\vee$ BL) $\wedge$ (EV)

Simplificando essa expressão e utilizando teoremas, propriedades e postulados da Álgebra Booleana, obtém-se a seguinte expressão minimizada: Fs(B) = (EV $\wedge$ (QP $\vee$ BL) $\wedge$ (EV $\vee$ QP $\vee$ BL)), que ainda pode ser escrita na forma de "Soma do Produto", ou seja, Fs(B) = (EV $\wedge$ (QP $\vee$ BL)). Os redutos são RED(B) = {Experiência do Vendedor, Qualidade do Produto} e {Experiência do Vendedor, Boa Localização}. A função de discernibilidade determinou o termo mínimo da função, ou seja, determinou o conjunto mínimo de atributos necessários para discernir quais as classes formadas por todas as classes de equivalência da relação INDs(B).

4. *Self-Organizing Maps* (rede SOM)

Um Mapa Auto-Organizável (*Self-Organizing Maps* ou rede SOM) é uma arquitetura de rede neural artificial com aprendizado não supervisionado, baseada em um mapa de neurônios cujos pesos são adaptados para verificar padrões semelhantes em relação a um conjunto de treinamento (KOHONEN, 2001). Sua principal característica é o mapeamento ordenado dos

padrões de entrada de elevada dimensão em reticulados de neurônios de saída com dimensão menor, comumente duas, o que facilita a visualização dos dados.

Para dada base de dados com N amostras com *d* atributos cada, em que *d* determina a dimensão dos padrões de entrada, ocorrerá mapeamento desses padrões para um reticulado de neurônios de saída arranjados em 2D, como mostrado na Figura 3.

A rede SOM é uma arquitetura de rede neural artificial, estruturada em duas camadas, entrada e saída. Os neurônios da camada de saída são comumente dispostos em um mapa de duas dimensões, com dada relação de vizinhança.

A Figura 3 ilustra essa arquitetura, com *d* atributos na camada de entrada e um conjunto de unidades *u* (neurônios) arranjados na forma de um mapa em 2D na camada de saída. Cada *u* é caracterizado por sua posição *x* e *y* no mapa, que é representado por *ux* e *uy*, respectivamente, resultando em um vetor 2D igual a *u =[ux uy]*. Cada *u* tem associado um vetor protótipo *mu = [m1u, m2u,..., mdu]*, sendo *d* a dimensão do protótipo, a mesma do padrão de entrada.

O algoritmo de aprendizado da rede SOM é realizado em um processo iterativo, em que no primeiro passo, t = 0, inicializa o vetor protótipo (m) randomicamente. Porém, a inicialização do m pode ser feita de outras maneiras (KOHONEN, 2001).

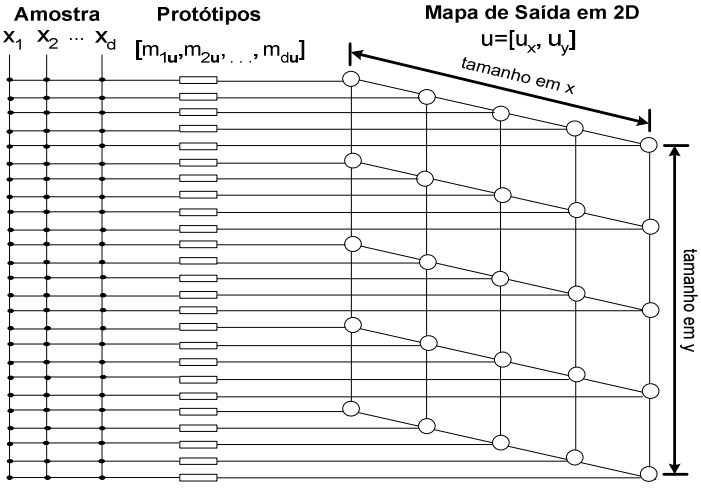

Figura 3 - Ilustração de uma rede SOM (2D)
Fonte: KOHONEN, 2001.

O algoritmo de treinamento da rede SOM é também chamado de competitivo. Em cada passo do processo (iteração ou época), uma amostra *x* é randomicamente escolhida do conjunto de treinamento. A distância, geralmente euclidiana, entre x e todos os vetores protótipos m é calculada. A unidade com menor distância, chamada de *best-matching unit* (BMU), é o *u* com protótipo m mais próximo de *x*, conforme a equação 1.

$$||x - mbmu|| = \arg \min_{u} || x-mu|| \qquad (1)$$

A seguir, os vetores protótipos são atualizados. O BMU e sua vizinhança topológica são movidos para próximos de *x*, como se fosse um "arrasto". A regra para a atualização dos vetores protótipos da unidade *u* é dada pela equação 2.

$$mi(t+1) = mi(t) + \alpha\,(t)\, h_{bi}(t)\, [x - mu(t)] \qquad (2)$$

em que *t* é o número de iterações, α (t) é a taxa de aprendizado e h_{bi} (t) é o *kernel* da vizinhança centrado no neurônio vencedor. O *kernel* pode ser gaussiano, como na equação 3.

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}} \qquad (3)$$

Em que r_b e r_i são as posições do neurônio vencedor *b* e do neurônio *i* no mapa da rede SOM, e σ(t) é o raio da vizinhança. Conforme a distância entre b e i aumenta e t também aumenta, $h_{bi} \rightarrow 0$. A taxa de aprendizado α(t) e o raio da vizinhança σ(t) diminuem monotomicamente com o tempo.

Devido às características da rede SOM de capacidade de quantização vetorial e de projeção vetorial, ele também pode ser utilizado na análise dos dados (KASKI; KOHONEN, 1996; CURRY et al., 2003). A quantização vetorial é feita com a projeção de N amostras de entrada para m protótipos, que representam todo o conjunto de dados original.

A partir dos protótipos, realizam-se a formação de grupos e a visualização das amostras em duas dimensões (JIN *et al.*, 2004). A Figura 4 traz como exemplo três grupos diferentes (dimensão d = 3).

O primeiro passo da rede SOM na análise de dados é a redução do conjunto de amostras para os m protótipos, os quais são utilizados no agrupamento ou na visualização dos dados.

A motivação para o uso dos protótipos é que a complexidade computacional do passo subsequente, a exemplo do agrupamento de dados, é reduzida. Quando utilizados na tarefa de *clustering*,

a rede SOM, em seu processamento das amostras, descarta ruídos com média zero e efeito de amostras discrepantes (*outliers*) na geração dos vetores protótipos (VESANTO; ALHONIEMI, 2000). Com os vetores protótipos, outros algoritmos de *clustering* podem ser utilizados, como a própria rede SOM ou o K-médias. Um estudo comparativo entre as redes SOM e K-médias é apresentado por Ultsch (1995).

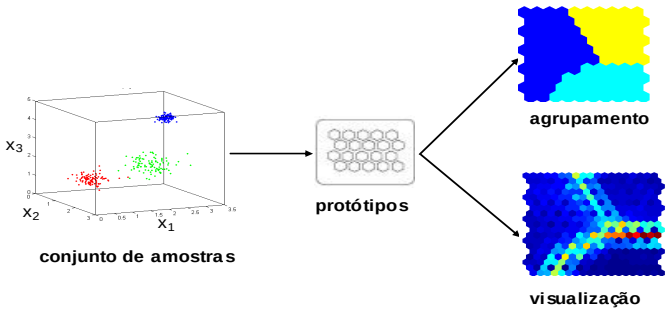

Figura 4 - Ilustração dos passos da utilização de uma rede SOM na análise de dados

5. Metodologia Experimental

O experimento realizado comparou a arquitetura híbrida proposta (RS com a rede SOM) com a rede SOM sem RS, analisando-se os resultados em termos de medidas de avaliação da rede SOM: erro de quantização (EQ), erro topográfico (ET), número de *clusters* gerados e visualização através do Mapa por Similaridade de Cor.

A escolha do tamanho do mapa foi baseada nas medidas EQ e ET. A melhor escolha é aquela que apresenta o menor EQ, que corresponde à média das distâncias entre cada vetor de dados ***x*** e o correspondente vetor de pesos do neurônio BMU e o menor ET, que quantifica a capacidade do mapa em representar a topologia dos dados de entrada. Para cada vetor de dados de entrada ***x*** são calculados seu primeiro BMU e o seu segundo BMU, e toda vez que eles não forem adjacentes (vizinhos, próximos) aumenta-se o erro em uma unidade, tirando, depois, a média pelo número total de vetores.

Esses indicadores expressam a capacidade do mapa em representar a topologia dos dados de entrada. Os menores valores desses erros, em geral, indicam melhor adaptação do mapa à topologia dos dados no espaço de entrada.

Os *clusters* são obtidos por intermédio da aplicação dos conceitos de similaridade e distância, e número maior ou menor de *clusters* pode indicar o que ocorreu com as similaridades entre os dados e as diferenças de separação dos *clusters*.

Cormack (CORMACK, 1971) afirmou que a tarefa de clusterização está pautada em duas ideias básicas, a coesão interna dos elementos e o isolamento externo entre os *clusters*. Dessa forma, um bom *cluster* é aquele que apresenta alta similaridade interna (intra-*cluster*) e baixa similaridade externa (extra-*cluster*).

A base de dados Consumidor (FERNANDEZ, 2003) é uma base que contém informações do consumo de 1.968 consumidores (registros) com 48 atributos, sendo 47 atributos condicionais e um atributo de decisão dividido em duas classes (M = masculino; F = feminino).

No caso do trabalho proposto, a base de dados escolhida não é extensa em número de registros, mas possui bom número de atributos (48), o que interessa ao RS, pois a técnica diminui atributos.

Foi eliminado o atributo que identificava o número da conta do consumidor. O atributo prefixo do nome é nominal e, portanto, foi transformado em numérico para poder ser processado pela rede SOM.

Os experimentos foram realizados em duas fases distintas descritas a seguir: na primeira fase denominada rede SOM sem redutos, apresentou-se à rede SOM a base de dados Consumidor com todos os atributos, e avaliaram-se os resultados.

Na segunda fase, denominada arquitetura híbrida (RS + rede SOM): apresentou-se primeiro ao RS a base de dados Consumidor para a geração dos redutos e, em seguida, a base de dados reduzida foi apresentada à rede SOM.

Para realização dos experimentos com a rede SOM foi utilizada a ferramenta SOM *Toolbox*, uma implementação do Mapa Auto-Organizável de Kohonen em Matlab. A escolha dessa ferramenta baseou-se no grande número de trabalhos publicados que relatam a sua utilização (VESANTO, 2000). De certa forma, essa ferramenta pode ser considerada como plataforma-padrão que tem sido adotada em grande parte das pesquisas atuais com Mapas Auto-Organizáveis. A *Toolbox* foi originalmente escrita para Matlab 5.0, no entanto funciona também nas versões mais atuais, de domínio público, podendo ser encontrada

para instalação no endereço <http://www.cis.hut.fi/projects/somtoolbox>.

Para a realização dos experimentos com RS foi utilizada a ferramenta chamada de Rosetta (*A Rough Sets Toolkit for Analisys of Data*), de domínio público, encontrada para instalação no endereço: <http://www.idi.ntnu.no/~aleks/rosetta>. Como acontece com a SOM *Toolbox*, a escolha da ferramenta Rosetta foi embasada no grande número de publicações que relatam a sua utilização (KOMOROWSKI; ØHM, 1997).

A plataforma de *hardware* utilizada nos experimentos foi um Pentium IV com 2.4 MHZ, 512 MB de memória RAM e 40 GB de disco rígido.

A escolha dos parâmetros, tanto de inicialização quanto de treinamento do mapa, ainda não segue regras bem definidas. Mesmo lançando mão das heurísticas existentes para definição dos parâmetros, é consenso entre os pesquisadores que se devem efetuar alguns testes com diferentes configurações de mapas antes de decidir qual representa melhor o conjunto de dados em questão. Em geral, são usadas heurísticas baseadas no comportamento do mapa e em medidas de qualidade como EQ e ET.

Os parâmetros que regulam a rede SOM podem ser agrupados em dois conjuntos: Parâmetros de estrutura - Dimensões (tamanho do mapa, número de neurônios), Vizinhança (hexagonal ou retangular) e Formato do Arranjo (folha, cilindro ou tiroide) - e Parâmetros de treinamento, que correspondem ao número de iterações (épocas) de treinamento.

Nos experimentos realizados, variaram-se alguns dos parâmetros de estrutura para conhecer a sua influência nos resultados. Dessa forma, todos os parâmetros descritos nesta seção foram utilizados nos experimentos tanto na primeira fase quanto na segunda, e são os seguintes:

Parâmetros de estrutura: Dimensões: número de neurônios 15X15 = 225 neurônios; Vizinhança: hexagonal (KOHONEN, 2001); e Formato do Arranjo: folha.

Parâmetros de treinamento: o treinamento de um mapa na SOM *Toolbox* é dividido em dois etapas: *rough phase* e *fine tune*. Em cada uma dessas etapas são definidos diferentes números de iterações. Os valores *default* (VESANTO, 2000) são: para a *rough phase* = 10 x mpd iterações; e para a *fine tune* = 40 x mpd iterações, em que mpd = neurônios/dados. A taxa de aprendizado foi de 0,5 na fase inicial e 0,05 na fase de convergência (KASKI; KOHONEM, 1996).O número de iterações foi calculado da seguinte forma: *rough phase* = quantidade de neurônios (225)/tamanho de dados (1968) x 10 = 12,0 e a *fine tune* = quantidade de neurônios (225)/tamanho de dados (1968) x 40 = 58,0. Somando as duas fases (12,0 + 46,0), têm-se 58 iterações. Após o treinamento do SOM, o mapa torna-se interessante ferramenta para a visualização dos dados (Figuras 5 e 6).

6. Resultados

Na primeira fase do experimento, apresentou-se à rede SOM a base de dados com todos os atributos condicionais (47), gerando 14 *clusters* (Figura 5). Na segunda fase, a base de dados Consumidor com todos os atributos condicionais (47) foi reduzida pelo RS, gerando 87 redutos, sendo 16 redutos com três atributos, 63 redutos com quatro atributos e oito redutos com cinco atributos.

Escolheram-se, então, os 16 redutos que apresentaram número menor de atributos (3) com base no menor esforço computacional (MITCHELL, 1997). Os 16 redutos apresentaram resultados muito semelhantes de EQ e de ET. Assim, com a ajuda do analista de dados o reduto escolhido foi o formado pelos seguintes atributos: valor da casa do consumidor, frequência de pedidos e idade do consumidor. Finalmente, a base de dados reduzida foi apresentada à rede SOM, gerando nove *clusters* (Figura 6).

A aplicação da arquitetura híbrida na base de dados Consumidor reduziu os valores de EQ e ET, indicando melhor representação da topologia da estrutura dos dados. Houve também redução no número de *clusters*, indicando melhor similaridade tanto entre os dados quanto entre as diferenças de separação dos agrupamentos. Essa diminuição no número de *clusters* e a melhor definição do mapaindicam que é uma boa base para, por exemplo, classificar os registros a respeito de certos critérios e que alguns consumidores, que eram considerados diferentes na primeira fase do experimento, na segunda fase foram agrupados em outros *clusters*, possibilitando maior coesão dos registros.

Os resultados levaram à conclusão de que o RS reduziu a informação que era apresentada à rede SOM, melhorando a formação dos *clusters*. Assim,

houve melhoria na visualização do mapa em razão da melhor definição das fronteiras (extra-*cluster*).

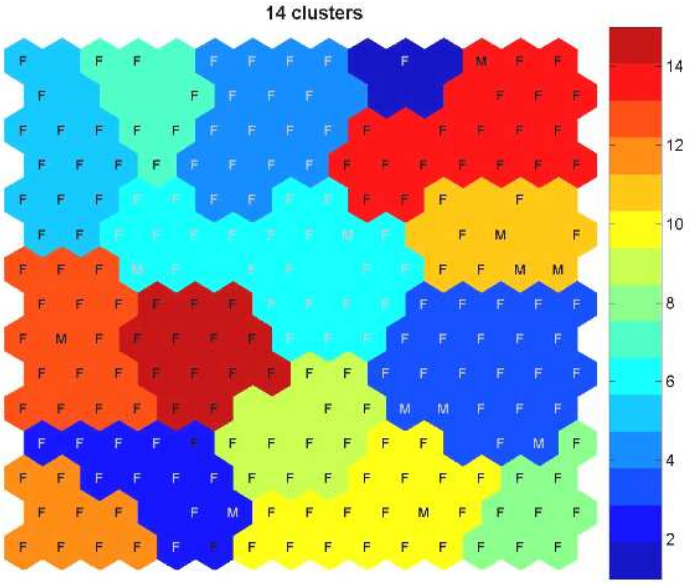

Figura 5 - Visualização dos 14 *clusters* gerados pela rede SOM sem redutos, rotulados de acordo com as classes da base de dados Consumidor (primeira fase do experimento)

A melhor definição dos *clusters* possibilita maior entendimento da base de dados, a fim de realizar atividades como: direcionamento nas campanhas de vendas, promoção de ofertas combinadas de serviços ou produtos, avaliação do comportamento do mercado e detecção de novas tendências mercadológicas ou necessidades de consumo.

O experimento possibilitou, também, maior conhecimento da base de dados, ou seja: sabe-se que o número de registros do sexo feminino é muito maior do que do sexo masculino, e existem perfis de consumo semelhantes entre os dois sexos. A existência de perfis de consumo semelhantes entre os dois sexos pode indicar mudança no comportamento de consumo ou nova tendência que está surgindo. Conclui-se que a arquitetura híbrida apresentou melhor desempenho do que a rede SOM sem RS.

Na Tabela 5, apresenta-se a comparação entre as duas fases do experimento.

Tabela 5 - Números resultantes da comparação entre a rede SOM sem redutos e a rede SOM com redutos (arquitetura híbrida)

	Número de *clusters*	**EQ**	**ET**	**Tempo**
SOM sem redutos	14	4,688	0,055	4s
Arquitetura híbrida	9	0,251	0,039	3s

Na Figura 6, procurou-se mostrar que a arquitetura híbrida por intermédio da rede SOM agrupou os registros da base de dados Consumidor obedecendo aos critérios de igualdade ou de semelhança entre os registros. A rede SOM utilizada no experimento é uma rede neural artificial 15 X 15, ou seja, possui 225 neurônios (Figura 6), contendo os registros dos consumidores agrupados em nove *clusters* pelo critério de igualdade ou semelhança. Para verificar se realmente a rede SOM (arquitetura híbrida) agrupou os consumidores segundo esses critérios, escolheram-se quatro neurônios no mapa (1, 15, 90 e 225, Figura 7).

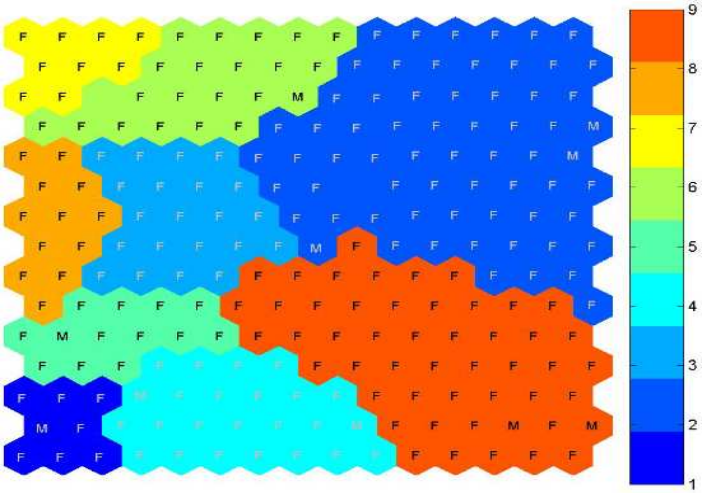

Figura 6 - Visualização dos nove *clusters* gerados pela redeSOM com redutos (arquitetura híbrida) rotulados de acordocom as classes da base de dados Consumidor (segunda fase do experimento)

A Tabela 6 mostra como a arquitetura híbrida agrupou os registros da base de dados Consumidor de acordo com os valores de cada atributo: valor da casa do consumidor, a frequência de pedidos e a idade do consumidor (reduto escolhido).

Tabela 6 - Informações dos atributos (reduto) considerados pela rede SOM para agrupar os neurônios

Neurônio	**Valor da casa**	**Frequência de pedidos**	**Idade**	**sexo**
1	98.700	4	43	F
15	**46.300**	**3**	**44**	**F**
90	**55.500**	**1**	**33**	**F**
225	142.900	2	29	F

Pode-se verificar, na Tabela 6, que os neurônios 15 e 90 (destacados em negrito) têm atributos com valores semelhantes e, por isso, pertencem ao mesmo *clust*er.

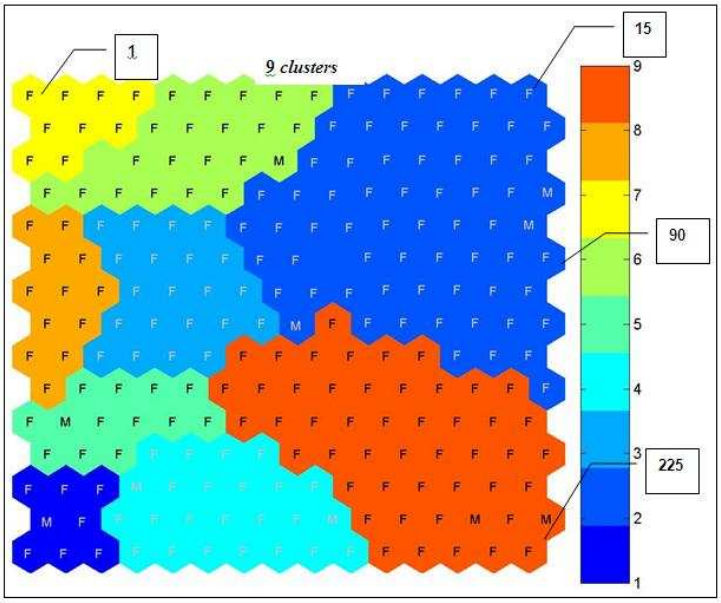

Figura 7 - Visualização dos neurônios localizados no mapagerados pela arquitetura híbrida.

Conclui-se que a rede SOM agrupou os registros obedecendo aos critérios de igualdade e semelhança de atributos (Figura 7). O neurônio 1 mostra que esse é um *cluster* de consumidores que possuem residência com valor intermediário, um cliente com número mais alto de pedidos e de meia-idade. Os neurônios 15 e 90 pertencentes ao mesmo *cluster* mostram que esse é um *cluster* de consumidores que possuem valor de residência mais baixo, frequência de pedidos num intervalo médio e de meia-idade.

O neurônio 225 mostra que esse é um *cluster* de consumidores que possuem casa com valor mais alto, frequência de pedidos média e são mais jovens. As conclusões a respeito de alto, médio e baixo valor da casa, frequência de pedidos alta, média e baixa e meia-idade, ou jovem, foram feitas com base na Tabela 6.

Percebe-se que a base de dados Consumidor possui número muito maior de casos do sexo feminino do que do sexo masculino. Isso justifica o fato de os *clusters* apresentarem número maior de letras F (feminino) do que M (masculino), e a colocação no mesmo *cluster* dos dois sexos evidenciam perfis de consumo iguais ou muito semelhantes, que podem ser verificados em maior profundidade utilizando classificador ou ferramentas de associação. No processo de KDD, essa fase também é conhecida como pós-processamento.

7. Conclusão

Como citado em Goldschmidt e Passos (2005), técnicas podem ser combinadas para gerar as chamadas arquiteturas híbridas. A grande vantagem desse tipo de sistema deve-se ao sinergismo obtido pela combinação de duas ou mais técnicas. Esse sinergismo resulta na obtenção de um sistema mais poderoso (em termos de interpretação, de aprendizado, de generalização, entre outros) e com menos deficiências.

Foi o que ocorreu com o desempenho da arquitetura híbrida proposta. Isso pode ser verificado na avaliação final dos resultados experimentais quando comparados com uma rede SOM sem a presença de RS, em que a arquitetura híbrida apresentou menor erro de quantização (EQ), menor erro topográfico (ET), menor número de *clusters* e melhor visualização do mapa gerado. Os menores valores de EQ e ET indicam que a arquitetura híbrida conseguiu representar a topologia dos dados de entrada melhor do que a rede SOM sem redutos.

Como os *clusters* são obtidos por intermédio da aplicação dos conceitos de similaridade e distância, número menor de *clusters* indica melhor similaridade entre os dados e maior diferença de separação dos *clusters*, ou seja, alta similaridade intra-*cluster* e baixa similaridade extra-*cluster*.

A redução de atributos realizada pelo RS fez que a informação considerada incerta não fosse apresentada à rede SOM, melhorando as fronteiras entre os *clusters*. Essa informação, quando submetida à rede SOM sem redutos, gerava incerteza, ocasionando em certos *clusters* definição de fronteira ruim, prejudicando a separação dos agrupamentos.

A combinação de RS com a rede SOM em uma arquitetura híbrida fez que uma das principais deficiências da rede SOM (a definição de fronteira entre os *clusters*) fosse melhorada, levando à conclusão de que em muitos casos é necessária a combinação de duas ou mais técnicas, a fim de eliminar ou reduzir certas deficiências individuais de cada técnica. Assim, com base nos experimentos e nos resultados, pode-se concluir que a arquitetura híbrida teve desempenho superior ao da rede SOM sem RS.

A redução da incerteza e a consequente melhoria na geração dos *clusters* obtida com a arquitetura híbrida possibilitam a formação de *clusters* mais bem definidos, pois os elementos da

base de dados que estavam na região de fronteira dos *clusters* foram agrupados melhor. Isso pode resultar numa geração de regras mais confiáveis por parte de um classificador ao ser utilizado após a submissão da base de dados à arquitetura híbrida.

Além das contribuições descritas, podem-se considerar aquelas do tipo como:

- Possibilitar maior conhecimento e maior difusão da Teoria dos RS ao revisar os principais conceitos da Teoria.
- Padronizar o confuso formalismo matemático da Teoria dos RS que foi encontrado na bibliografia pesquisada.
- Demonstrar a aplicação dos principais conceitos da Teoria dos RS na diminuição da incerteza contida em uma base de dados.
- Apresentar outra opção para o tratamento da incerteza, além das já tradicionalmente conhecidas como a Teoria dos Fuzzy Sets e a Distribuição de Probabilidade em Estatística.
- Demonstrar que a rede SOM é uma ótima técnica para minerar dados, pois possibilita em um mapa bidimensional a formação e a visualização simples dos clusters e da correlação dos dados, preservando a posição relativa desses clusters no hiperespaço original.

Finalmente, com base nos resultados, pode-se considerar que a aplicação da arquitetura híbrida pode ser vantajosa em diversas áreas-alvo de KDD, como: *marketing* (detecção de perfil do consumidor), medicina (imagens), governamental (detecção de fraudes), financeiras (concessão de crédito), entre outras.

Affonso e Sassi (2010), em trabalho na área da Engenharia de Produção aplicando arquitetura híbrida *Rough*-Neuro *Fuzzy* ao processamento de polímeros, obtiveram bons resultados.

Os estudos aqui realizados não têm a pretensão de esgotar o assunto; pelo contrário, buscou-se proporcionar contribuição com o desenvolvimento da arquitetura híbrida para descobrir conhecimento em bases de dados. Sabe-se que há clara demanda por estudos sistematizados que estabeleçam outros domínios de aplicação ainda mais adequados para a arquitetura híbrida proposta. Este cenário oferece, portanto, amplo espaço para trabalhos de continuidade.

8. Referências

AFFONSO, C.; SASSI, R. J. An inference mechanism for polymer processing using Rough-Neuro Fuzzy Network. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS ICANN 2010, part III, LNCS 6354. **Proceedings...** Springer, 2010. p. 441-50.

BERRY, M. J. A.; LINOFF, G. **Data mining techniques**: for marketing, sales and customer Support. [S.l.]: John Wiley & Sons, 1997.

CASTANHEIRA, L. G. Aplicação da mineração de dados à análise das condições de operação de transformadores. **Revista Eletrônica Produção e Engenharia**, v. 2, n.1,p. 12-3, jan./jul. 2009.

CORMACK, R. M. In: A review of classifications. **Journal of Royal Statistical Society**, Series A, 134, 1971; p. 321-67, 1971.

CURRY, B.; DAVIES, F.; EVANS, M.; MOUTINHO, L.; PHILLIPS, P. The Kohonen Self-organizing Map as an alternative to cluster analysis: an application to direct marketing". **International Journal of Market Research,** v. 45, Quarter 2, jun. 2003.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. In the KDD process for extracting useful knowledge from volumes of data. **Comunications of the ACM**, v. 39, n.11, p. 27-34, 1996.

FERNANDEZ, G. **Data mining using SAS applications**. New York: Chapman & Hall/CRC, 2003.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining um guia prático**. Conceitos, técnicas, ferramentas, orientações e aplicações. Rio de Janeiro: Campus, 2005.

HAN, J.; KAMBER, M. **Data mining**: concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2001.

HAYKIN, S. **Neural networks**: a comprehensive foundation. New York: Willey & Sons, 1999.

JIN, H.; SHUM, W. H.; LEUNG, K. S.; WONG, M. L. Expanding self-organizing map for data

visualization and cluster analysis. **Information Sciences**, v. 163, p. 157-73, 2004.

KASKI, S.; KOHONEN, T. Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In: INTERNATIONAL CONFERENCE ON NEURAL NETWORKS IN THE CAPITAL MARKETS, 3., 1996. **Proceedings…** [S.l.], 1996. p. 498-507.

KIVILUOTO, K. Topology preservation in self-organizing maps. In: INTERNATIONAL CONFERENCE ON NEURAL NETWORKS (ICNN'96), 1996, New York. **Proceedings…** New York, 1996. v. 1, p. 294-9.

KOHONEN, T. **Self-organizing maps**. Springer series in information sciences. 3. ed. New York: Springer; Berlin, 2001. v. 30.

KOMOROWSKI, J.; ØHRN, A. ROSETTA: a rough set toolkit for analysis of data. In: INTERNATIONAL JOINT CONFERENCE ON INFORMATION SCIENCES, 3.; INTERNATIONAL WORKSHOP ON ROUGH SETS AND SOFT COMPUTING (RSSC'97), 5., 1997, Durhan. **Proceedings…** Durham, NC, USA, mar. 1-5, v. 3, p. 403-7, 1997.

KUMAR, U. A.; DHAMIJA, Y. Comparative analysis of SOM neural network with K-means clustering algorithm. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF INNOVATION AND TECHNOLOGY (ICMIT) –IEEE, 2010. **Proceedings…** [S.l.], 2010. p. 55-9.

LABIOD, L.; GROZAVU, N.; BENNANI, Y. Relational topological clustering. In: THE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2010. **Proceedings…** [S.l.],2010. p. 1- 8.

MITCHELL, T. **Machine learning**. New York: Mcgraw-Hill, 1997.

MOHEBI, E.; SAP, M. Rough set based clustering of the self-organizing map. In: ASIAN CONFERENCE on INTELLIGENT INFORMATION AND DATABASE SYSTEMS (ACIIDS), 1., 2010. **Proceedings…** [S.l.], 2010. p. 82-5.

MONGIOVI, G. **Uso de relevância semântica na melhoria da qualidade dos resultados gerados pelos métodos indutivos de aquisição de conhecimento a partir de exemplos**. 1995. Tese (Doutorado) – Universidade Federal da Paraíba, João Pessoa, 1995.

PAL, S. K.; DASGUPTA, B.; MITRA, P. Rough self-organizing map. **Applied intelligence (Special issue on Soft Case Based Reasoning)**, v. 21, p. 289-99, 2004.

PATRÍCIO, C. M. M. M.; PINTO, J. O. P.; SOUZA, C. C. Rough sets: técnica de redução de atributos e geração de regras para classificação de dados.In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL – CNMAC**,** 28., 2005. **Proceedings…** [S.l.], 2005.

PAWLAK, Z. Rough sets. In: INTERNATIONAL JOURNAL OF COMPUTER AND INFORMATION SCIENCES, 11., 1982. **Proceedings…** [S.l.],1982. p. 341-56.

PAWLAK, Z. **Rough sets**: theoretical aspects of reasoning about data. Kluwer, 1991.

SANT'ANNA, A. P. Rough sets analysis with antisymmetric and intransitive attributes: classification of brazilian soccer clubs. **Pesquisa Operacional**, 28, n. 2, p. 217-30, 2008.

SASSI, R. J.; SILVA, L. A.; DEL MORAL HERNANDEZ, E. A Methodology using Neural Networks to Cluster Validity Discovered from a marketing Database. In: BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS (SBRN), 10., 2008. **Proceedings...** [S.l.], 2008. p. 3-8.

UCHÔA, J. Q. **Representação e indução de conhecimento usando teoria de conjuntos aproximados.** 1998. Dissertação (Mestrado) – Universidade Federal de São Carlos, São Carlos, SP, 1998.

ULTSCH, A. Self-organizing neural networks perform different from statistical K-means Clustering. **Proceedings of GfKl**. Basel, Swiss, 1995.

VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. **IEEE Transaction on Neural Network**, v.11, p. 586-600, 2000.

Recebido em 19.04.2010
Publicado em 29.01.2011